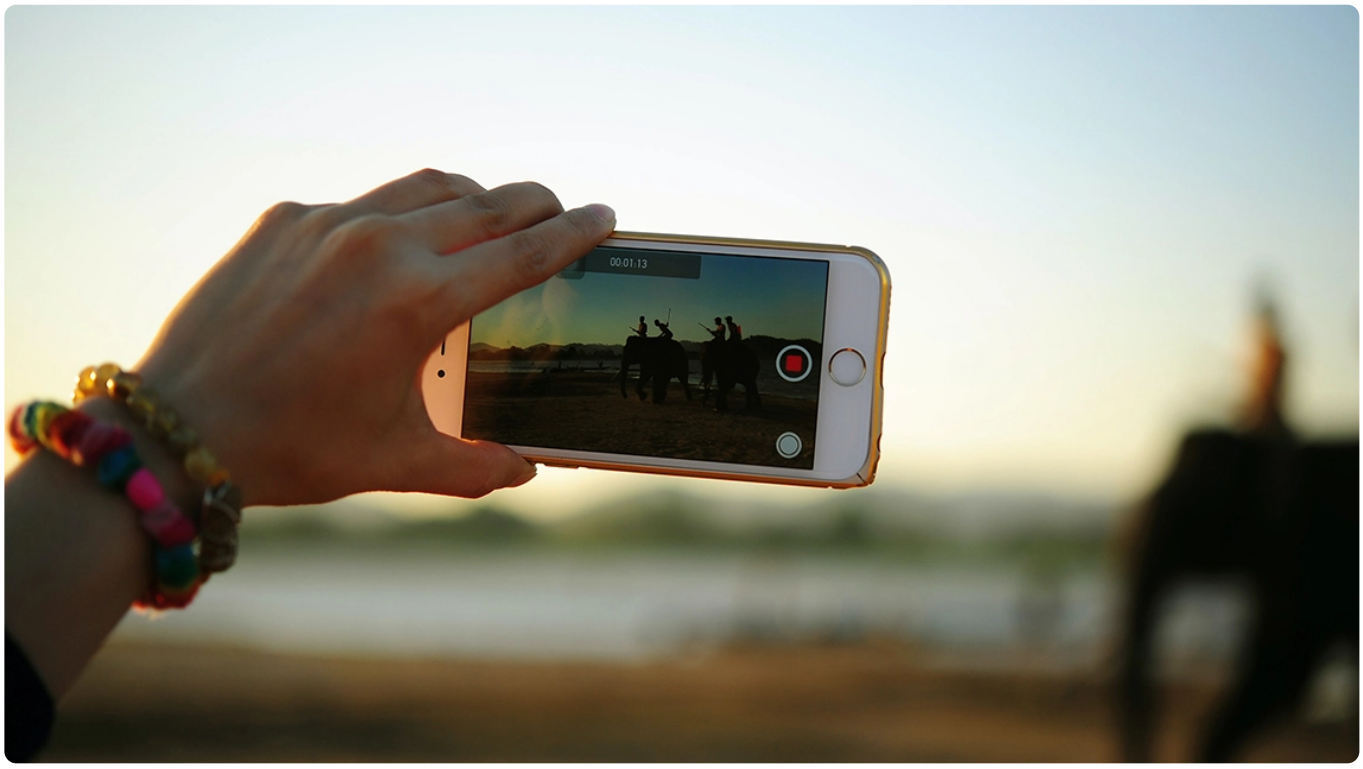
01｜基本概念：从参数的角度看视频图像

2021-11-22 李江

《攻克视频技术》 课程介绍 >



讲述：李江
时长 13:30 大小 12.38M

你好，我是李江。

从今天开始，我们会一起来学习一些视频和图像相关的技术。主要包括视频图像的基本概念、图像的缩放处理、视频压缩编码、视频打包传输以及音视频同步等相关知识。

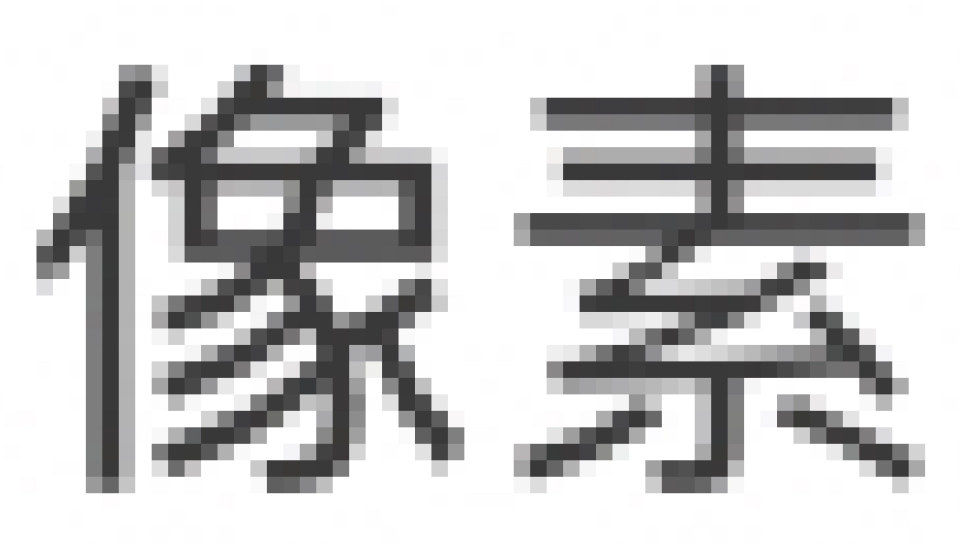
今天，我们就从视频和图像（视频是由一帧帧图像组成的）的基础知识讲起。掌握了这些之后，我们再讨论如何对图像进行缩放、如何使缩放后的图像更加清晰，以及如何对视频进行编码压缩等就更加游刃有余了。当然了，这些话题更难，但也更有意思，希望我们能有个不错的开始！

像素

相信你对像素这个概念一点都不陌生。从智能手机市场大火到现在，我们经常能够听到某某最新款手机，多少多少万像素。像素越高，则图像就会越清晰，拍出来的图片就会更逼真。那像素到底是什么呢？

像素是图像的基本单元，一个个像素就组成了图像。你可以认为像素就是图像中的一个点。

我们来直观地看看像素是怎么组成图像的。在下面这张图中，你可以看到一个个方块，这些方块就是像素。



那一张图片有多少个像素呢？要回答这个问题就需要引出另外一个非常重要的概念——分辨率。

分辨率

图像（或视频）的分辨率是指图像的大小或尺寸。我们一般用像素个数来表示图像的尺寸。比如说一张 1920x1080 的图像，前者 1920 指的是该图像的宽度方向上有 1920 个像素点，而后者 1080 指的是图像的高度方向上有 1080 个像素点。

视频行业常见的分辨率有 QCIF（176x144）、CIF（352x288）、D1（704x576 或 720x576），还有我们比较熟悉的 360P（640x360）、720P（1280x720）、1080P（1920x1080）、4K（3840x2160）、8K（7680x4320）等。

那么**同样一张图像用不同的分辨率表示会有什么不同呢？**我们可以通过以下这组图片来直观感受一下。

我们可以看到，1x1 的时候我们只能看到一个像素，只有一种颜色，根本就不是我们想象中的图像了。而 10x10 的图像几乎都是糊的，只能看到一点点轮廓。随着图像的分辨率越来越高，图像的细节就越来越清晰。由此，我们可以总结出：

1. **像素就只是一个带有颜色的小块。**
2. **图像的分辨率越高，图像就越清晰。**

但从更加专业的角度来说，第 2 句话不够严谨。原始图像的话，分辨率越高确实会越清晰，但是我们看到的图像往往是经过后期处理的，比如放大缩小，或者磨皮美颜。经过处理过后的图像，尤其是放大之后的图像，分辨率很高，但是它并没有很清晰。

这是因为放大的图像是通过“插值”处理得到的，而插值的像素是使用邻近像素经过插值算法计算得到的，跟实际相机拍摄的像素是不一样的，相当于“脑补”出来的像素值。

因此，放大的图像还是会存在偏差，表现出来就是会模糊。我们会在之后的课程中来具体聊聊这个过程是怎么做的。总之，**我们不能简单地认为分辨率数值越高的图像就越清晰。**

刚才我们在前面还提到，像素就是一个带有颜色的小块，那这个小块到底是怎么组成的呢？这里我们就来讲讲 RGB 图像像素和位深的概念。

位深

一般来说，我们看到的彩色图像中，都有三个通道，这三个通道就是 R、G、B 通道。简单来说就是，彩色图像中的像素是有三个颜色值的，分别是红、绿、蓝三个值。也就是说我们看到的那个带有颜色的块其实是由 R、G、B 三个值组成的（有的时候还会有 Alpha 值，代表透明度，我们这里不展开讨论）。

通常 R、G、B 各占 8 个位，也就是一个字节。8 个位能表示 256 种颜色值，那 3 个通道的话就是 256 的 3 次方个颜色值，总共是 1677 万种颜色。我们称这种图像是 8bit 图像，而这个 8bit 就是位深。

我们可以看到，**位深越大，我们能够表示的颜色值就越多。**因此，图像就可以更精确地展示你拍摄的真实世界。比如现在有 10bit 图像和 12bit 图像，8bit 图像的每一个像素需要占用 3x8 总共 24 个位，3 个字节，同理 10bit、12bit 就会占用更多。

所以，图像的位深越大，需要的存储空间就会越大，传输这张图像使用的流量就会越多。目前我们大多数情况下看到的图像以及视频还是 8bit 位深的。

Stride

接下来我们来看一个特别的概念——Stride。这个 Stride 不是图像本身的属性，但是视频开发者经常会碰到，也是经常会出问题的一个东西。我们团队在工作中就多次遇到过由于客户没有处理好这个东西，从而导致播放的图像出现“花屏”的情况。

Stride 也可以称之为跨距，是图像存储的时候有的一个概念。**它指的是图像存储时内存中每行像素所占用的空间。**你可能会问，一张图像的分辨率确定了，那一行的像素值不就确定了吗？为什么还需要跨距这个东西呢？其实，为了能够快速读取一行像素，我们一般会对内存中的图像实现内存对齐，比如 16 字节对齐。

举个例子，我们现在有一张 RGB 图像，分辨率是 1278x720。我们将它存储在内存当中，一行像素需要 1278x3=3834 个字节，3834 除以 16 无法整除。因此，没有 16 字节对齐。所以如果需要对齐的话，我们需要在 3834 个字节后面填充 6 个字节，也就是 3840 个字节做 16 字节对齐，这样这幅图像的 Stride 就是 3840 了。如下图所示：

这个地方你一定要注意，**每读取一行数据的时候需要跳过这多余的 6 个字节**。如果没有跳过的话，这 6 个字节的像素就会被我们误认为是下一行开始的 2 个像素（每个像素 R、G、B 各占 1 个字节，2 个像素共 6 个字节）。那这样得到的图像就完全错了，显示出来的就是“花屏”现象，屏幕会出现一条条的斜线。

所以，不管你去读取还是渲染一张图片，还是说你将这张图片存储下来，都需要设置正确的 Stride。很多时候，尤其是不规则分辨率的时候，它和图像的 Width（R、G、B 的话就是 Width x 3）是不一样的。

有的时候即便图像的 Width 是一个规则的值，比如说 1920 或者 1280 等能被 16 整除的宽度，图像存储在内存中有可能 Stride 和 Width（R、G、B 的话就是 Width x 3）也是不一样的，尤其是不同的视频解码器内部实现的不同，会导致输出的图像的 Stride 不一样。

所以，一定要在处理图片的时候注意这个 Stride 值。如果出现一条条斜线的花屏或者说解码后图像的颜色不对的情况，我们需要先确认一下这个 Stride 值对不对。

帧率

以上就是图像的基本概念，接下来我们来讲讲视频的一些基本概念。

前面我们说到，视频是由一系列图像组成的，即“连续”的一帧帧图像就可以组成视频。但事实上，视频中的图像并不是真正意义上的连续。也就是说，在 1 秒钟之内，图像的数量是有限的。只是当数量达到一定值之后，人的眼睛的灵敏度就察觉不出来了，看起来就是连续的视频了。

这个 **1 秒钟内图像的数量就是帧率**。据研究表明，一般帧率达到 10～12 帧每秒，人眼就会认为是流畅的了。当然，可能会有个体差异。

通常，我们在电影院看的电影帧率一般是 24fps（帧每秒），监控行业常用 25fps，而我们声网常用的帧率有 15fps、24fps 和 30fps。你可以根据自己的使用场景来具体设定你想使用的帧率值。

选择帧率的时候还需要考虑设备处理性能的问题，尤其是实时视频通话场景。**帧率高，代表着每秒钟处理的图像数量会很高，从而需要的设备性能就比较高**。如果是含有多个图像处理过程，比如人脸识别、美颜等算法的时候，就更需要考虑帧率大小和设备性能的问题。同样，也要考虑带宽流量的问题。帧率越大，流量也会越多，对带宽的要求也会越高。

码率

我们已经知道，视频的帧率越高，1 秒钟内的图像数据量就会越大。通常我们存储视频的时候需要对图像进行压缩之后再存储，否则视频会非常大。

那么压缩之后的视频我们一般如何描述它的大小呢？一般对于一个视频文件，我们直接看视频的大小就可以了。但是在实时通信或者直播的时候，视频是视频流的形式，我们怎么衡量呢？

这就涉及到我接下来要介绍的概念——码率。码率是指视频在单位时间内的数据量的大小，一般是 1 秒钟内的数据量，其单位一般是 Kb/s 或者 Mb/s。通常，**我们用压缩工具压缩同一个原始视频的时候，码率越高，图像的失真就会越小，视频画面就会越清晰**。但同时，码率越高，存储时占用的内存空间就会越大，传输时使用的流量就会越多。

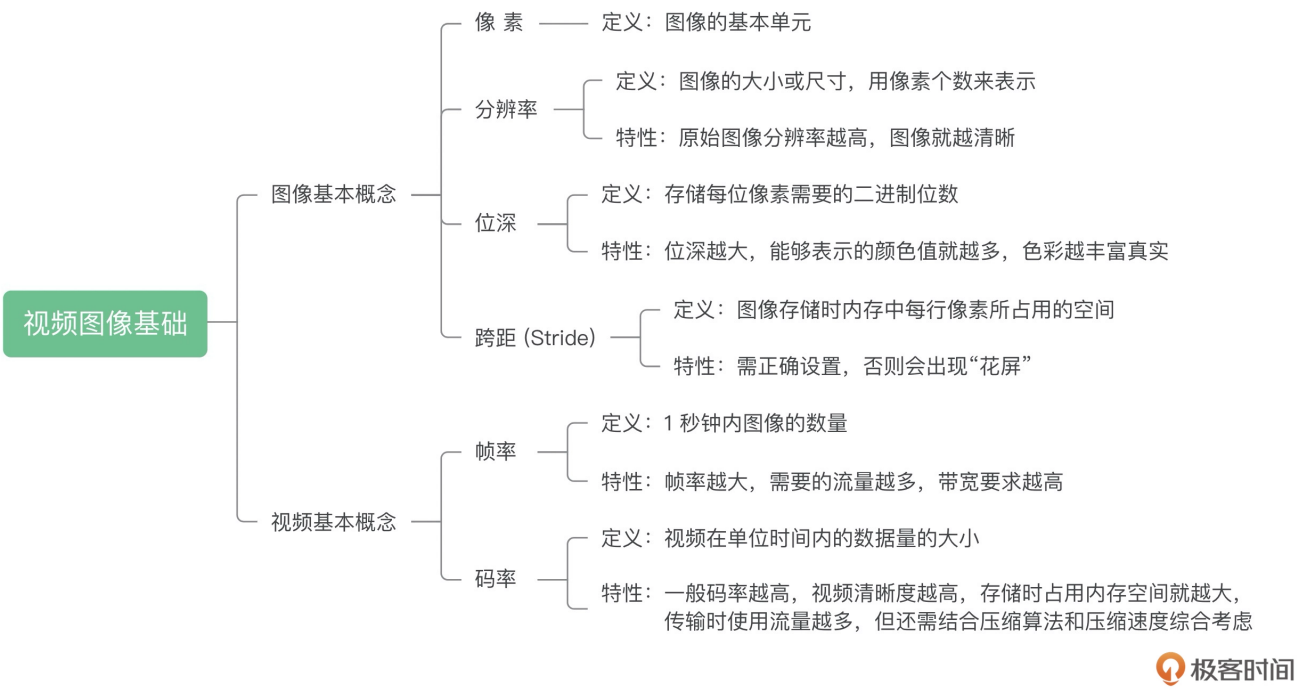
这里请你思考一个问题，同一个原始视频被压缩之后，真的是码率越高，清晰度就越高吗？

其实准确来说的话，不是。因为视频的压缩是一个非常复杂的过程，之后我们会有好几节课来讲视频压缩的知识。事实上，视频压缩之后的清晰度还跟压缩时选用的压缩算法，以及压缩时使用的压缩速度有关。压缩算法越先进，压缩率就会越高，码率自然就会越小。压缩速度越慢，压缩的时候压缩算法就会越精细，最后压缩率也会有提高，相同的清晰度码率也会更小。

所以，**并不是码率越高，清晰度就会越高。**

小结

今天我们学习了图像和视频的基础知识，都很简单但很重要，这里我为你总结了一张图帮助你记忆。



总结来说，一张图像是由像素组成的，而图像有多少像素则由分辨率来表示。在分辨率之外，存取一副图像还需要特别注意 Stride 这个东西，它跟分辨率中的 Width 是不一样的。然后，一帧帧图像组成了视频，我们将每秒中的图像数量称之为帧率。视频编码后每秒的数据量称之为码率。

这些知识点是我们之后课程的基础，随着我们不断深入学习，还会不断巩固这些概念。

思考题

现在请你想一想：码率是固定的，还是会变化的？如果是固定的，怎么做到呢？

这节课到这里就结束了，欢迎留言和我分享你的思考和疑惑，你也可以把今天所学分享给身边的朋友，邀请他加入探讨，共同进步。

分享给需要的人，Ta订阅后你可得 20 元现金奖励

生成海报并分享

赞 6 提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 开篇词 | 为什么说视频开发是程序员的一片蓝海？

下一篇 02 | YUV & RGB：原来图像是这么丰富多彩的

精选留言 (14)

写留言

西格玛

2021-11-22

stride比较难理解，小白问几个stride的问题：
1.stride为什么是16字节对齐？
2.为什么stride是宽度16字节对齐，高度却不限制？
3.读取图片的时候，如果就把stride当成宽度（实际宽度小于stride），还会花屏吗？（stride大于宽度的部分我当成纯色）

展开

作者回复: 1、一般芯片读取效率在16字节对齐时最高，芯片能一次性读取16字节的数据，16字节对齐之后可以充分利用芯片的性能，速度更快。当然并不是stride一定是16字节对齐的，有的也32字节对齐，甚至64字节对齐。
2、图像存储在内存中一般是个一维数组，一般我们需要一行一行像素读取，不太会一列一列读

取。因此高度方向上一般不对齐。

3、不会花屏，但是会有多余的不是图像的部分，一般不会留着这个padding部分。

共 3 条评论 > 4

贾献华

2021-11-23

码率可以是固定的，也可以是变化的。

如果是固定码率，

1. 编码后的码率小于固定码率，填充数据
2. 编码后的码率大于固定码率，丢弃细节数据，降低码率

展开

作者回复: 是的。回答的很对。一般来说除非要求绝对固定，不然不会填充数据。毕竟浪费带宽。

你的回答考虑地很全面。

3

DK

2021-11-23

码率可以是固定的,但是更倾向于使用变化的

首先我们储存视频时并不是以图片的形式进行存储,而是以图片中不同色值的变化作为一个存储的依据,在不同的时间点内,运动色值的轨迹变化量可能不同,所以最终的结果导致了不同,考虑的最优算法

如果不能确定色值轨迹的变化量,而是固定的存储某个位置的像素点变化,码率就是固定的

展开

作者回复: 一般来说：

1、对清晰度要求高，可以选择变化的码率。

2、RTC场景，一般使用固定码率。

1

Geek_00444b

2021-11-23

码率可以变化也可以恒定，码率如果是固定的，可以通过一个缓冲区来暂时存储码率，从而使输出的码率恒定，但是加了缓冲区会增加时延

展开

作者回复: 这个思路有点像webrtc中的paced_sender的思路。但是一般来说我们还是会使用提高QP和降低QP的方式来实现码率基本恒定。提高QP画面会变模糊一些，会有可能出现马赛克，但是码率会降低。降低QP画面会变清晰。码率也会提高。

共 3 条评论 > 2

哦吼掉了

2021-11-23

1.图像由像素组成,有多少像素取决于分辨率,对原始图像来说分辨率和清晰度成正相关。图像以一定帧率组成视频,视频以一定的码率进行传输。
2.基本概念
位深:一幅图像中包含的二进制位的数量
位深越大,能够表示的颜色值就越多。RGB通道。1位深,只能表示黑白;...
展开

作者回复: 2、位深是一个像素点中某一个通道（比如R通道、G通道、B通道）的像素值的二进制位的数量。

共 2 条评论 > 2

_zeta

2021-11-22

老师讲的太好了，声网服务这么好还是因为有大神坐镇的。
刚开始做音视频不久，有个问题想请教，我们在iOS平台自采集后用 VideoToolBox 进行H.264编码，写入SEI时，我们自研的播放器是可以解析SEI的，但问题是用ffplay拉流会报[h264 @ 0x14786b800] nal_unit_type: 1(Coded slice of a non-IDR picture), nal_ref_idc: 1...
展开

作者回复: 看起来是在没有IDR之前就出了非IDR帧，导致第一帧解码不了。

共 2 条评论 > 2

plh

2021-11-30

小白用户问个问题，码率是怎么计算出来的。
展开

作者回复: 对于播放器来说，实时码率一般是一秒钟多少帧，从编码好的码流中读取这个多帧数据，就知道总共多少字节大小了，码率就是总字节数除以一秒。对于实时发送的视频流一般你可以知道你一秒钟发送了多少字节的，那码率也就知道了。对于编码器来说它一秒钟输出多少肯定也知道，很容易算码率。对于普通人看一个视频文件的码率。可以直接用文件总大小除以总时长，但是一般视频文件里面还有音频数据，所以得到的是视频和音频的总码率，通常音频的码率相比视频小很多，可以近似认为是视频的码率。

Alan
2021-11-29

可以客观评价下，目前国内的音视频服务厂商整体指标的对比么？最近在选择更合适的音视频服务商，希望老师给个参考

展开

作者回复: 实时音视频体验可以从音视频延时、卡顿、登陆成功率、出图时间、音质、画质等几个维度进行综合评估，并需要关注一下哪些厂商宣传和实际的是否是一致的，最好是能提供质量保证。

1

Geek_ba06dc
2021-11-26

思维导图用什么软件u做的

展开

作者回复: 我一般用xmind。我觉得还挺好用的。

yu
2021-11-24

4K分辨率的文稿写错了吧，语音读的应该是对，4096×2160

共 3 条评论 >

1

麋鹿少年
2021-11-24

大佬，之后会讲讲webrtc的码率自适应吗

展开

作者回复: 会详细介绍。

共 2 条评论 1

ShawnL

2021-11-24

在压缩图像以后，视频里面的图片像素还能保持原来的尺寸么(e.g 1080p)? 尤其是如果为了固定码率或者减小码率的情况下，是否可以理解为视频的分辨率就要降低。

作者回复: 会保持分辨率，减小码率的方式可以关注视频压缩的课程，就在下一周会更新。

Geek_828b39

2021-11-24

D1分辨率不是704×576或者720×576么？

展开

作者回复: 是的，笔误了，多谢兄弟指出错误。

共 2 条评论

西钾钾

2021-11-23

码率应该是编码器控制的既可以是固定的也可以是变化的。
1、变化我理解是，视频的内容不同时，就会产生变化，比如视频的背景特别简单和特别复杂这两种情况下编码出来的码率就是不一样的。
2、固定我理解是，比如上面同样的场景在背景简单和复杂的情况下，编码器在处理复杂背景是压缩的狠，在处理简单背景是不压缩。...

展开

作者回复: 回答的很好，很厉害。但是简单背景我们还是会压缩的，只是压缩的失真很小。

1